Some generic measures of the extent of chemical disequilibrium applied to living and abiotic systems

B. F. Intoy and J. W. Halley
School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455
(Dated: August 22, 2018)

We report results of evaluation of several measures of chemical disequilibrium in living and abiotic systems. The previously defined measures include R_T and R_L which are Euclidean distances of a coarse grained polymer length distribution from two different chemical equilibrium states associated with equilibration to an external temperature bath and with isolated equilibration to a distribution determined by the bond energy of the system, respectively. The determination uses a simplified model of the energetics of the constituent molecules introduced earlier. We evaluated the measures for data from the ribosome of E. Coli, a variety of yeast, the proteomes (with certain assumptions) of a large family of prokaryotes, for mass spectrometric data from the atmosphere the Saturn satellite Titan and for commercial copolymers. We find with surprising consistency that R_L is much smaller than R_T for all these systems. Small R_L may be characteristic of systems in the biosphere.

I. INTRODUCTION

To define the mission of the search for extraterrestrial biology with quantitative precision, it would be desirable to specify measures of molecular systems which are more generic than searches for particular molecules or combinations of molecules, such as those in the earth's biosphere. That is because it is not known that all nonequilibrium systems having lifelike characteristics will turn out to be at all chemically similar to the terrestrial biosphere. Designing such measures is a goal of our program of study of Kauffman-like models of prebiotic evolution. The choice of such measures constitutes an implicit definition of the meaning of the term 'lifelike'. However if the measures are quantitatively well defined, then one can use them without engaging in a debate about whether certain ranges of their values imply , necessarily or sufficiently, a 'lifelike' system.

Here we evaluate several such generic measures for a set of experimentally known systems, including some which are in the biosphere and others which are not. A goal is to determine how well the measures distinguish living from nonliving systems and what they reveal about the differences. The measures used here were first applied by us to the results of simulations of a Kauffman-like model as described in [1]. However we did not need to assume a detailed correspondence between that model and the experimentally observed systems described here in order to carry out the program reported in this paper. We do need to assume that the systems contain only linear polymers and we make some simplifying assumptions about the bonding energies as described in more detail below. In particular, the assumptions about the energy allot an upper bound to the energy, and thus allow equilibrium with negative temperatures as discussed below. These assumptions are well satisfied for some of the systems under study and for others they are more problematical as we will discuss.

The main experimental inputs to the analysis presented here are the molecular weight density distributions, the numbers of monomers available for forming polymers (parameter b defined below) and two parameters characterizing the size of a polymer coil of L monomers. The main outputs include values of two dimensionless parameters termed R_T and R_L which describe how far the observed system is from chemical equilibrium with an external thermal bath and from chemical equilibrium if isolated, respectively. Detailed definitions are provided in section II.

The systems we study are the proteomes of 4,555 prokaryotes, the ribosomes in E. coli, a budding yeast, five artificial copolymers, and the atmosphere of Titan. The choice of systems was made in order to provide a sampling of living systems and some engineered system (the copolymers). Titan's atmosphere was selected because Titan's atmosphere is reported [2] to contain an abundance of organic molecules and may be characteristic of a prebiotic environment [4].

We find some interesting trends: We find that all the systems are closer to isolated equilibrium with a high temperature which is negative for the copolymer and Titan systems than they are to chemical equilibrium with the (positive) ambient temperature of their surroundings. The living systems have a sharply defined R_T value different from the values found for the other systems. It appears that a small R_L and a large well defined R_T distinguish the living from the nonliving systems. This feature is closely related to the fact that in peptide systems under biological conditions the bonding energy as defined in our model is negative whereas it is positive for the other systems. We present detailed molecular weight distributions to contrast what is observed experimentally with the corresponding equilibria.

In the next section we define the essential features and assumptions which allow definition and evaluation of the measures R_L , R_T and other generic characteristics. We focus first on the case in which all the bond energies are the same followed by a brief discussion of the case in which there is more than one bond energy. Section III reports the processes by which we extracted the needed information from the raw experimental data in each of

the cases considered. In section IV we report the results and section V contains a discussion and conclusions.

II. ANALYSIS

We first describe the case in which all the polymeric bond energies are the same, as we have assumed for all systems except the Titan atmosphere. Polymers are assumed to consist of strings of monomers with b types possible for each monomer. We took $b = 20$ for the ribosomes in E. coli, the prokaryotes and for yeast and $b = 2$ for the Titan data and the copolymer data. To any polymer of length L we attribute an energy $-(L-1)\Delta$ where Δ is a real number which is the bonding energy between two monomers. It is taken to be positive for the copolymers and the Titan atmosphere but negative for the peptide bonds in the biological systems as discussed in the next section. The total energy E of any population $\{n_m\}$ of polymers in which n_m is the number of polymers of type m is $E = -\sum_{L=1}^{l_{max}}(L-1)N_L\Delta$. Here the $N_L = \sum_{m\ of\ length\ L} n_m$ is the same set of macrovariables used in [5] and [6]. We denote the total number of polymers N in a sample by is $N = \sum_{L=1}^{l_{max}} N_L$. However, in contrast to the situation in the dynamic simulations described in [1], the input data for calculation of equilibrium distributions are not N and E but the volumetric polymer concentration $\rho = N/V$ where V is the solution volume and the volumetric energy density $e = E/V$. To take entropic account of the dilution of the experimental sample we introduce a microscopic length $R_0 L^\nu$ where R_0 is a length related to the polymer persistence length and ν is an index which would be 1/2 for a random walk. We report the numerical values for ν and R_0 used for the various systems considered in the next section. We modify the expression for the entropy used in [1] to take account of the number of ways to distribute N polymers in a volume V as follows: $S/k_B = \ln W$ with

$$W = \prod_L \frac{(N_L + G_L - 1)!}{N_L!(G_L - 1)!}$$

and $G_L = b^L V/v_p L^{3\nu}$ and $v_p = R_0^3$. b is the number of monomers available for inclusion in the polymers in the system and ν is a dimensionless number which is 1/2 for a random walk. The expression is identical to the one used in [1] except for the factor $V/v_p L^{3\nu}$ in the degeneracy G_L. Note that a similar configurational entropy factor was used by us in an earlier paper [6]. With this modification we apply Stirling's approximation and maximize the entropy subject to the density and energy constraints associated with the experimental data discussed in the next section. We have

$$S/k_B = \sum_L [\ln((G_L + N_L - 1)!) - \ln(N_L!) - \ln((G_L - 1)!)] \tag{1}$$

Proceeding in the standard way to maximize the entropy under these constraints we have when both energy density $e = E/V$ and polymer number density $\rho = N/V$ are fixed that the values $\overline{N_L}$ of the populations which maximize this entropy are

$$\overline{N_L} = \frac{G_L - 1}{\exp(-\beta(e,\rho)\mu(e,\rho) - \beta(e,\rho)\Delta(L-1)) - 1} \tag{2}$$

Here the parameters $\beta(e,\rho)$ and $\mu(e,\rho)$ are determined from the total energy density e and polymer number density $\rho = N/V$ by the implicit equations (with (2))

$$e = -(1/V)\sum_{L=1}^{l_{max}}(L-1)\overline{N_L}\Delta \tag{3}$$

and

$$\rho = (1/V)\sum_{L=1}^{l_{max}}\overline{N_L} \tag{4}$$

We use the definition of G_L and define $v_p = R_0^3$ to write these relations as

$$\overline{N_L(v_p/V)} = \frac{b^L/L^{3\nu} - (v_p/V)}{\exp(-\beta(e,\rho)\mu(e,\rho) - \beta(e,\rho)\Delta(L-1)) - 1} \tag{5}$$

and

$$ev_p = -\sum_{L=1}^{l_{max}}(L-1)\overline{N_L(v_p/V)}\Delta \tag{6}$$

and

$$\rho v_p = \sum_{L=1}^{l_{max}}\overline{N_L(v_p/V)} \tag{7}$$

These are in dimensionless form, convenient for solving for $\beta(e,\rho)\mu(e,\rho)$ and $\beta(e,\rho)\Delta$ numerically because they do not involve macroscopically large numbers. The term v_p/V on the right hand side of (5) is in all cases much less than $b^L/L^{3\nu}$ and is dropped in the numerical analysis. As before [1] we refer to this equilbrium as 'isolated'. There is no reference to an external temperature bath. Also, as in [1] we also determine a 'thermal' equilibrium distribution by solving (7) with a fixed value of β using reported approximate values of the ambient temperature in the experiments considered and making no use of e. In each case we can then use (5) to evaluate the polymer length density distributions expected in those two equilibrium states.

Because the systems of interest are not necessarily (and in fact are found not to be) in either kind of chemical equilibrium we used the experimentally observed values of the quantities $N_L(v_p/V)$ to evaluate Euclidean distance in the space of values of sets $\{N_L v_p/V\}$ between the actual population set $\{N_L v_p/V\}$ and the ones corresponding to the two kinds of equilibria given by (5)

with $\beta(e,\rho)$, $\mu(e,\rho)$ in the isolated case and with a fixed ambient β and $\mu(\rho)$ in the case in which the system is equilibrated to an external bath. Thus we define two Euclidean distances R_L and R_T in the l_{max} dimensional space of sets $\{N_L(v_p/V)\}$ which characterize how far the system of interest is from the two kinds of equilbria:

$$R_L = \sqrt{\sum_L (v_p/V)^2 (N_L - \overline{N_L(\beta(e,\rho),\mu(e,\rho))})^2}/(\sqrt{2}v_p\rho) \tag{8}$$

for distance from the locally equilibrated state and

$$R_T = \sqrt{\sum_L (v_p/V)^2 (N_L - \overline{N_L(\beta,\mu(\beta,\rho))})^2}/(\sqrt{2}v_p\rho) \tag{9}$$

for distance from the thermally equilibrated state.

In the cases of prokaryotes, E. coli ribosomes, yeast and the copolymers which we evaluate, we use the assumptions as just described. However in the analysis of the data from the Titan atmosphere, we take account of the large difference between the energies of CC and CN bonds and NN bonds by reformulating the description with two bond strengths as described in Appendix A and reference [9]. We show there that an accurate treatment of that case gives results very close to those obtained by using the same model as the one described above, but with an average bond strength $\overline{\Delta}$ of

$$\overline{\Delta}(p_c) = \Delta_{CC} M p_c^2 + 2\Delta_{CN} p_c(1-p_c) + \Delta_{NN}(1-p_c)^2. \tag{10}$$

p_c is the atomic fraction of the atmosphere which is carbon. (We assume $\Delta_{CC} = \Delta_{CN}$ as discussed further in the next section and Appendix B).

III. EXTRACTION OF POPULATION DISTRIBUTIONS FROM DATA

Here we describe how data was extracted from data available on a budding yeast, 4,555 prokaryotes, the ribosomes in E. Coli, the commercial copolymers and mass spectrographic data on the atmosphere of Titan in order to compute the characteristic parameters $R_L, R_T, \beta\mu$ and $\beta\Delta$ defined in the preceding section. We report a comparison of the results in the next section in order to determine the extent to which these parameters may distinguish living from lifelike or nonliving systems.

A. Yeast

We used data on protein population distributions in the budding yeast *Saccharomyces cerevisiae* from reference [16] with[7] $R_0 = 1.927$Åand $\nu = 0.588$ in solving equations (6) and (7) . For determining the chemical potential in the case of equilibrium with an external temperature bath (equation (7)) we used a temperature of 293 K and a bond energy [8] of -2.2 kcal/mol. It is important to note that for this and the other protein systems Δ is taken to be negative, meaning that it costs energy to make a peptide bond. The value of -2.2 kcal/mol taken from reference [8] is the Gibbs free energy of reaction which can be seen from the description of the model in [1] to be the appropriate energy to identify with Δ within our model.

B. Ribosomes in E. coli

We used protein distribution data from [20]. There are approximately 50000 ribosomes per cubic micron in an E. coli cell [21] the ribosome giving a value for Nv_p/V of about 1.86×10^{-5} assuming that, in equilibrium, the proteins would be denatured and evenly distributed throughout the cell. We took $R_0 = 1.927$Åand $\nu = 0.588$ as for the yeast and prokaryotes.

C. Prokaryote Proteins

We used data from the Kyoto Encyclopedia of Genes and Genomes[14] (KEGG) (the faa , fasta amino acid file) to extract approximate length distributions for 4,555 prokaryotes assuming that there is only one of each protein in each prokaryotic cell. The KEGG data gave amino acid sequences for all the proteins, from which the coarse grained number N_L of all proteins of each length L was extracted. To get a volumetric distribution N_L/V by averaging data in [15] we obtained an approximate average volumetric protein density of 2.5 million proteins per cubic micron which was used for all the prokaryotes when solving equations 7 and 6. As for yeast, the form[7] $v_L = R_0^3 L^{3\nu}$ with $R_0 = 1.927\mathring{A}$ and $\nu = 0.588$ was used in the solutions to equations 7 and 6 . For thermal equilibrium calculations a temperature of 20 degrees Celsius and a bond energy of -2.2 kcal/mol was used.

D. Copolymers

We used data from references [17, 18]. The experimental details are summarized in Table I. The data references given in Table I give the relative abundances of polymers as a function of their composition. The table also mentions the monomers used in the systems which were either polystyrene-polyisoprene or polystyrene -polybutadiene. For example the data from [17] gives the relative abundances $f(l_s, l_i)$ of polymers in a solution of a polystyrene-polyisoprene system as a function of their styrene l_s and isoprene l_i compositions, as extracted from mass spectrometry data. To get values of f_L we sum all the values of $f(l_s, l_i)$ for which $l_s + l_i = L$. From [17], the total number of monomer blocks in the copolymer system was 8.6 mol, consisting of 0.7 mol of styrene and 7.9 mol of isoprene. From the length distribution $\{f_L\}$the average

polymer length is calculated to be approximately 28.9. Dividing the total number of monomer blocks by the average polymer length gives the total number of polymers to be approximately $N = 1.79 \times 19^{23}$. The experiment took place in a $V = 6$ Liter reactor vessel giving N/V. For solving equations (6) and (7) we obtained v_p for this system using listed values for the persistence lengths l_p, statistical segment lengths $\tilde{b}$ and characteristic ratios C_∞ cited in [25] to obtain the value $R_0 = \tilde{b}^{1-\nu} l_{eff}^{\nu}$ with $l_{eff} = 2l_p/\sqrt{C_\infty}$ the effective bond length as defined in [25]. For the copolymers we assumed that $\nu = 0.5$ and used the average of the R_0 values calculated from the parameters b, l_p and C_∞ for Isoprene and Styrene in [25]. For calculations of the chemical potential μ from (7) when the system is in thermal equilibrium with an external thermal bath, a bath temperature of 293 K and the average Carbon-Carbon bond energy of 347 kJ/mol was used [19].

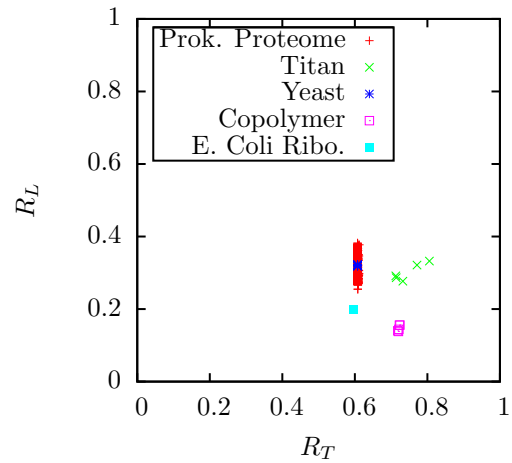

FIG. 1: R_T and R_L summary scatter plot for all the systems considered.

E. Titan Data Extraction and Analysis

We used mass spectrographic data on the atmosphere of Titan taken by the *Cassini* spacecraft as reported in [2]. The data on mass distributions is reported in [2] in units of charge detected per second. In this paper we only report analysis of the Cassini mass spectroscopy data on detected negative ions. The mass distribution of negative ions contains the widest mass distribution reported, including molecules as massive a 10^4 daltons. Data are available for detected neutral and positively charged molecules [3] and may be analysed later. We assume in our analysis that all the molecules detected had unit charge(in units of the magnitude of the electron charge). Masses were reported in daltons and converted approximately to monomer units by dividing by an assumed average monomer mass of 13 daltons because the monomers are believed to be predominantly single nitrogen or carbon entities. (We are neglecting the contribution of the hydrogen masses). We converted the data reported in [2] in log scale bins to linear scale bins as described in Appendix A. To extract volumetric densities N/V we used the kinetic relation $\dot{N} = \epsilon A(N/V)v$ where $\dot{N}$ is the rate at which particles are detected per per second, $\epsilon = 0.05$ is the detector efficiency [2], A is the detector area, reported to be 0.33 cm^2 [2] and v is the velocity of the spacecraft relative to the Titan atmosphere for which we used $v = 6.3 km/sec$ from reference [10]. This procedure gave volume densities for data taken at different altitudes above the Titan surface as summarized in Table III. We assumed that the species detected in the mass spectrometer were singly charged. Uncertainties resulting from this procedure are discussed in [2]. For the value of $v_p = R_0^3$ in equations (6) and (7) we used the $R_0 = 4.97$ Å extracted from data for polyethylene in [25] in the same way that it was done for the copolymers as explained above (with $\nu = 0.5$).

For determining the equilibrium distribution in the presence of an ambient thermal bath using equation (7) we used an ambient temperature of 120 degrees Kelvin [12].

In applying equations (7) and (6) to the Titan data, we are assuming that all the molecules detected are linear chains, which is certainly not expected to be true [4]. Also, CC and CN bond energies are quite closely similar as described in Table II, but the N-N bond is much weaker. In principle, one should therefore take account of the different bond energies in the model used to obtain equations (6) and (7) which we solve to determine the equilibrium distributions. We have done that, with results described elsewhere [9]. However it turned out that a simplifying 'mean-field' approximation works quite well to describe the results and we use that here: We used an 'average' bond energy $\Delta = \overline{\Delta}$ where

$$\overline{\Delta}(p_c) = \Delta_{CC} p_c^2 + 2\Delta_{CN} p_c(1-p_c) + \Delta_{NN}(1-p_c)^2. \quad (11)$$

with $\Delta_{CC} = \Delta_{CN} = 325 kJ/mol$ and $\Delta_{NN} = 160 kJ/mol$. p_c is the atomic fraction of the atmosphere which is carbon. We use the value $p_c = 0.02$ [13] . As explained in Appendix B, we also obtained results using the maximum and minimum values which the bond strength could have at the observed p_c and found that the results were quite insensitive to the change and consistent with the results of [9].

IV. RESULTS

We summarize the data found for R_T and R_L in Figure 1 with a second view focussing on the region where the data cluster in Figure 2.

The yeast distribution is shown in Figure 3. The distributions for three of the 4,555 prokaryotes analysed are shown in Figures 4-6 and the E. coli ribosomes length

TABLE I: Parameters used for the five copolymer systems considered here. $N_{\rm mon}$ is the total number of monomer blocks in units of mols, $\langle L \rangle$ is the average polymer length, and V is the volume in liters. R_0 values are calculated by taking the average of the two monomer types from [25].

Data Reference	Monomer types	Figure in this paper	$N_{\rm mon}$ (mols)	$\langle L \rangle$	V (Liters)	R_0(Å)	Density (Nv_p/V)
Figure 7 of [17]	Isoprene and Styrene	8	8.6	28.9	6	5.10	0.00395
Figure 6-13b of [18]	Isoprene and Styrene	9	1.21	32.9	2	5.10	0.00147
Figure 6-13c of [18]	Isoprene and Styrene	10	1.21	41.3	2	5.10	0.00117
Figure 6-13d of [18]	Isoprene and Styrene	11	1.21	49.9	2	5.10	0.000968
Figure 6-15 Left of [18]	Styrene and Butadiene	12	0.966	36.6	2	5.27	0.00116

TABLE II: The average bond energies for carbon and nitrogen [19].

Bond	Average Bond Energy (kJ/mol)
C-C	347
C-N	305
N-N	160

TABLE III: Titan atmospheric densities for various altitudes as detected during the 40th Titan encounter of Cassini which occurred on 05 January 2008.

Altitude (km)	Figure in this paper	Density (Nv_p/V)
1013	13	4.56×10^{-20}
1032	14	3.79×10^{-20}
1078	15	1.72×10^{-20}
1148	16	6.54×10^{-21}
1244	17	7.89×10^{-22}

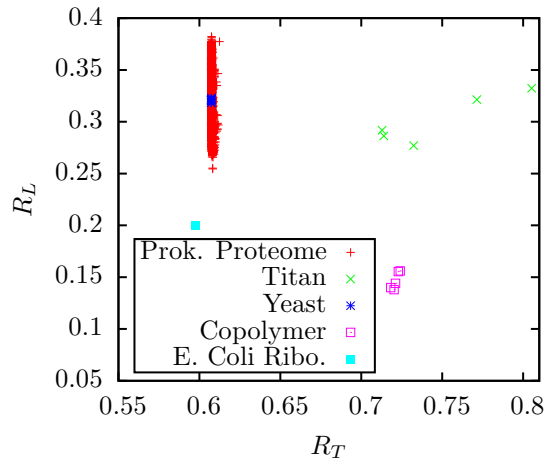

FIG. 2: Data from the preceding figure (Fig. 1) showing only the region where system points appear.

distribution appears in Figure 7. Length distributions for five copolymer systems are shown in Figures 8-12. Figures 13-17 compare the observed length distributions with the calculated equilibrium for the Titan data at five elevations. Values of R_L and R_T for the data exhibited in the figures appears in Table IV.

In figure 18 we show a scatter plot of the average polymer length versus the standard deviation of the length distribution for the same set of systems. The next figure 19 shows the values of $\beta\Delta$ and $\beta\mu$ found for the isolated equilibrium state in each case.

V. DISCUSSION AND CONCLUSIONS

The results show quite definitely that the prokaryotic and yeast proteomes are far from chemical equilibrium with the ambient environment as measured by nearly identical values of about $R_T \approx 0.61$ found for all of them. The reason for this is that the large number (b=20) of available monomers (amino acids) plays a significant role in determining the thermal equilibrium population distribution. (For all cases considered here, the terms -1 in the numerator and denominator of equation 2 can be ignored (the 'Gibbs limit').) The dominant factor determining the equilibrium distribution is $e^{(\ln b+\beta\Delta)L}$. When $\Delta < 0$ as it is for polypeptides the exponent can be negative. If the exponent is positive the thermal equilibrium is predominantly long polymers whereas if it is negative, the distribution is predominantly short polymers. This is a manifestation of the competition between entropy, which in the case of peptides drives the system toward long polymers, and energy, which drives it toward monomers. When $b = 20$ the external temperature , (or the bond energy) must be finely tuned to avoid one of these extremes and for most reasonable values of Δ and β the fine tuning does not occur and the thermal equilibrium state is described by one of the extremes. The R_T value is then the distance in the space of coarse grained populations between one of those extreme cases (all monomers or all polymers of maximum length) and the actual, nonequilibrium population distribution. Because the actual polymer distributions of the living systems are all similar and the thermal equilibrium distribution is at one of the extremes, the R_T values tend to be the same. However because of the sensitivity of the thermal equilibrium population to the value of $\beta\Delta$ one finds one or the other of the two extremes with values of Δ which are within the range of measured peptide bond energies. We illustrate

TABLE IV: Values of parameters found here characterising the observed population distributions of the systems considered. (Equations 7 and 6 were solved with six figure precision but only three significant figures are reported here.)

Figure in this paper	Description	R_L	Local $\beta\mu$	Local $\beta\Delta$	R_T	Thermal $\beta\mu$	Thermal $\beta\Delta$
3	Yeast	0.319	−15.2	−2.99	0.607	−14.5	−3.78
4	Prokaryote baw	0.254	−15.0	−2.99	0.608	−14.1	−3.78
5	Prokaryote cch	0.382	−14.6	−3.00	0.608	−14.1	−3.78
6	Prokaryote cva	0.333	−14.7	−2.99	0.607	−14.1	−3.78
7	E. coli Ribosomes	0.199	−15.4	−2.98	0.597	−14.1	−3.78
8	Copolymer JCA 2010	0.156	−8.57	−0.570	0.724	-6.44×10^3	142
9	Copolymer Staal 13b	0.155	−9.53	−0.586	0.723	-7.45×10^3	142
10	Copolymer Staal 13c	0.144	−9.96	−0.597	0.721	-8.88×10^3	142
11	Copolymer Staal 13d	0.138	−10.3	−0.605	0.720	-1.03×10^4	142
12	Copolymer Staal 15	0.140	−9.52	−0.607	0.718	-8.88×10^3	142
13	Titan 1013km	0.292	−45.8	−0.693	0.713	-1.68×10^5	167
14	Titan 1032km	0.286	−45.9	−0.694	0.714	-1.68×10^5	167
15	Titan 1078km	0.277	−46.7	−0.697	0.732	-1.68×10^5	167
16	Titan 1148km	0.321	−47.6	−0.708	0.771	-1.68×10^5	167
17	Titan 1244km	0.332	−49.8	−0.694	0.805	-1.68×10^5	167

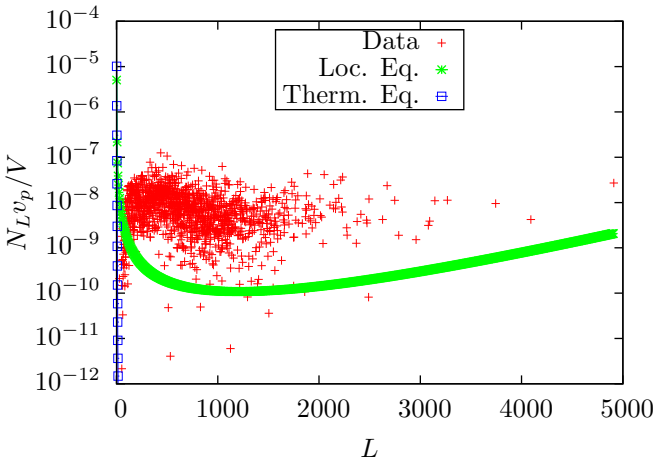

FIG. 3: $N_L v_p/V$ values for yeast with glucose as its carbon source together with the corresponding local and thermal equilibrium distributions which are respectively euclidean distances (see equations (8) and (9)) $R_L = 0.32$ and $R_T = 0.61$ away from the observed distribution.

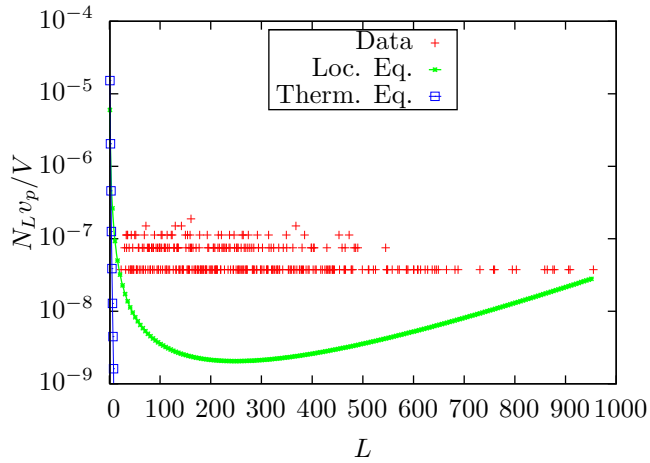

FIG. 4: $N_L v_p/V$ values for the proteins found in the prokaryote Buchnera aphidicola JF98 an endosymbiont of Acyrthosiphon pisum which has organism code baw on KEGG and corresponding equilibrium distributions. $R_L = 0.25$ and $R_T = 0.61$.

this in Figure 20 where we show data for a prokaryote analysed using the average (-2.2 kcal/mol) of the values of the peptide bond energy reported in reference [8] and the results of the same analysis using the average plus one standard deviation and the average minus one deviation. For the least negative value of $\beta\Delta$ the thermal distribution jumps from all monomers to all polymers of maximum length.

We conclude that within our model, with ambient temperatures in the range of 100's of degrees, a system of polypeptides with a large b (which is 20 for terrestrial biochemistry) will be very sensitive to the external temperature and will be driven toward long polymers at higher temperatures and monomers at lower ones with a sharp transition in between. The exact transition will be determined by the value of Δ (< 0). In the real systems, that value of the peptide bond energy varies from one amino acid pair to another and the corresponding sharp transition will be smeared by the distribution of bond energies. We illustrate these points in another way in Figure 21 where we show the calculated value of R_T for a series of $\beta\Delta$ values for one of the data sets. There is a sharp minimum in R_T when the thermal distribution is at the tipping point between all monomers and maximum

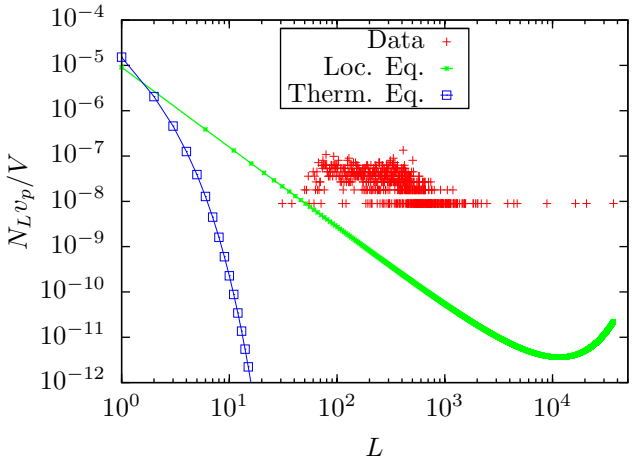

FIG. 5: $N_L v_p/V$ values for the proteins found in the prokaryote Chlorobium chlorochromatii CaD3 which has organism code cch on KEGG and and corresponding equilibrium distributions.. $R_L = 0.38$ and $R_T = 0.61$.

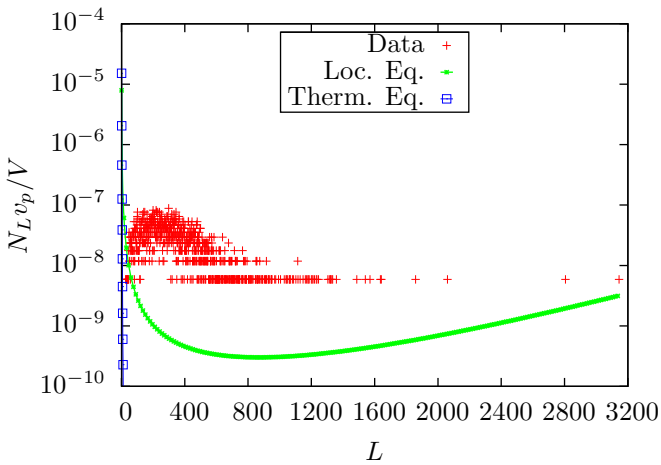

FIG. 6: $N_L v_p/V$ values for the proteins found in the prokaryote Corynebacterium variable DSM 44702 which has organism code cva on KEGG and corresponding equilibrium distributions. $R_L = 0.33$ and $R_T = 0.61$.

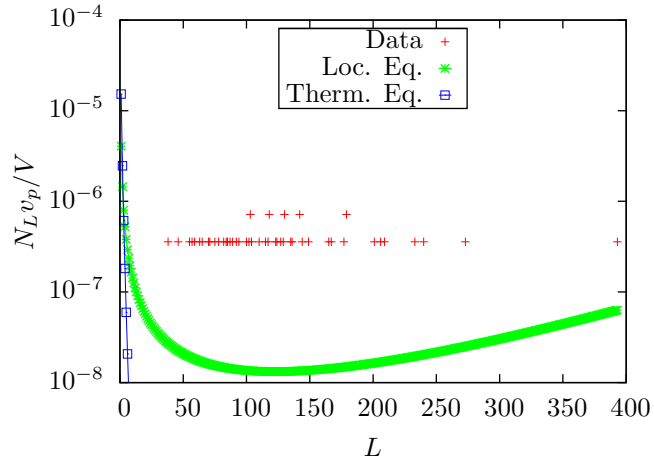

FIG. 7: $N_L v_p/V$ values for the proteins in the E Coli ribosome .and corresponding equilibrium distributions. $R_L = 0.20$ and $R_T = 0.60$.

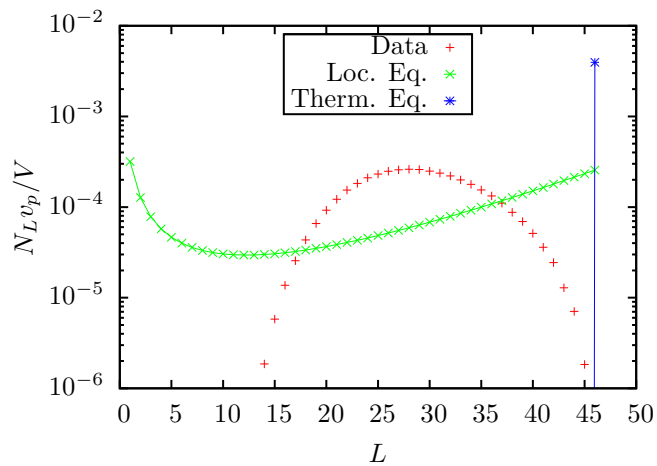

FIG. 8: $N_L v_p/V$ values for the data from Figure 7 of [17] and corresponding equilibrium distributions‘. $R_L = 0.16$ and $R_T = 0.72$.

length polymers. The minimum is close to but different from the local value of $\beta\Delta$ which was calculated from the protein energy and polymer number.

The implication seems to be that the observed protein distributions in the living systems are quite close to a local distribution which is quite unlikely since it requires a fine tuned value of $\beta\Delta$ to be produced in a thermal environment. One may construct the following plausibility argument to speculatively explain this: Generally, a large number of possible chemical configurations must be explored in order to find those special ones which result in lifelike properties such as autocatalysis. If a system is in a thermal environment which strongly favors all monomers or all long polymers, not many configurations will be explored. Hence the systems from which a lifelike system is most likely to emerge will be those which are thermally fine tuned to be in the intermediate region in which the system has a spread of polymer lengths and, we expect, large fluctuations in the polymer length distribution. Once autocatalysis begins, the system may become self stabilizing and its local distribution will differ from the one imposed by its thermal environment, as we observe for the prokaryotes, the yeast and the ribosome. Hence we suggest that the 'local' chemical equilibrium associated with the polymer density and bond energy density of these living systems, which is quite close to the actual population distributions (though the values of R_L which measure this disequilibrium vary between about 0.25 and about 0.38) might be a relic of the early thermal environment in which those living systems orig-

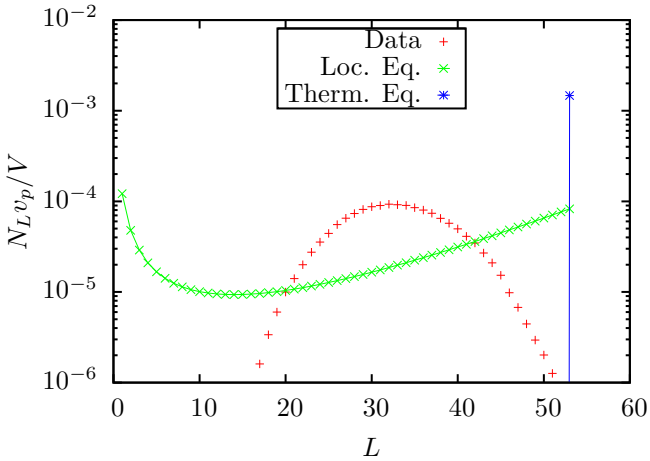

FIG. 9: $N_L v_p/V$ values for the copolymer system for which taken data are given in Figure 6-13b of [18] and corresponding equilibrium distributions.. $R_L = 0.16$ and $R_T = 0.72$.

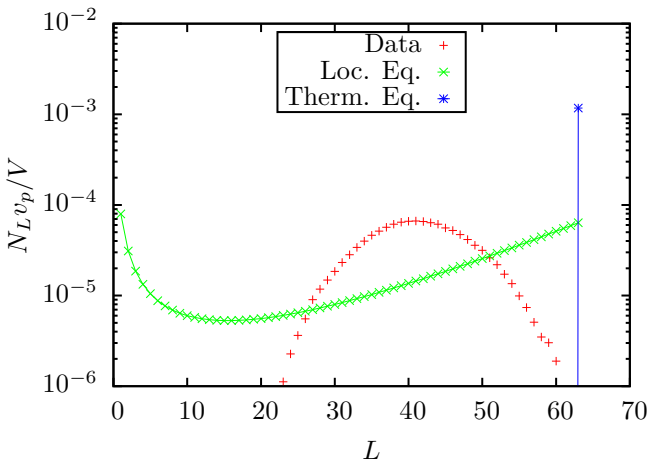

FIG. 10: $N_L v_p/V$ values for the copolymer system for which data are given in Figure 6-13c of [18] and corresponding equilibrium distributions. $R_L = 0.14$ and $R_T = 0.72$.

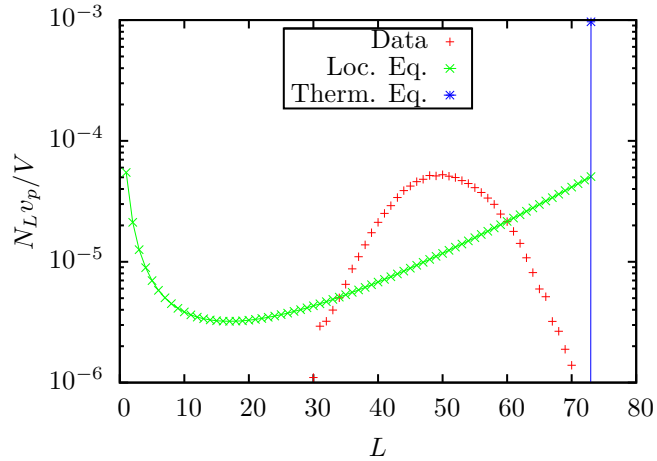

FIG. 11: $N_L v_p/V$ values for the copolymer system taken for which data are given in Figure 6-13d of [18] and corresponding equilibrium distributions. $R_L = 0.14$ and $R_T = 0.72$.

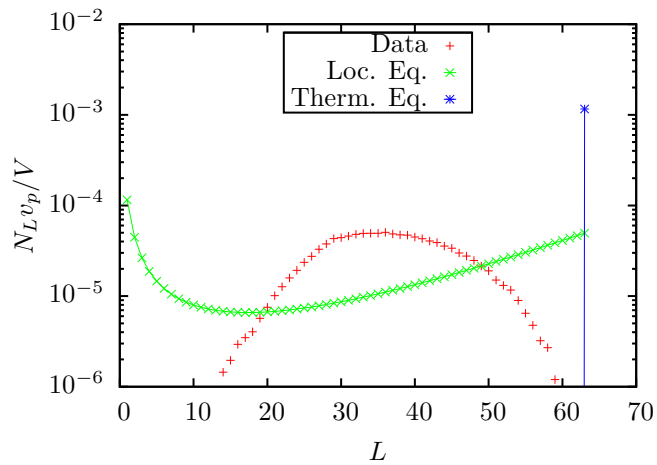

FIG. 12: $N_L v_p/V$ values for the copolymer system for which data are given in Figure 7-15 Left of [18]i and corresponding equilibrium distributions. $R_L = 0.14$ and $R_T = 0.72$.

inally evolved. The needed value of temperature is easily estimated from the relation $\ln b = -\beta\Delta$ from which $T = |\Delta|/(k_B \ln b)$. With the average bond energy we have used and $b = 20$ this is within a few degrees of the boiling point of water at 1 atmosphere.

These speculations suggest experiments with peptide systems in which the value of the effective $\beta\Delta$ is varied over a very fine scale to find the tipping point and discover if any interesting behavior emerges near it. For a more careful estimate of the needed temperature one could take the expected (or measured) temperature dependence of the bond energy into account and solve the resulting implicit equation for T.

The living systems have values of the parameters R_T and R_L which are robustly distinct from those for the copolymer ($R_T \approx 0.725, R_L \approx 0.15$) and the atmosphere of Titan $0.71 < R_T < 0.82, 0.28 < R_L < 0.33$. In these two measures, Titan is mainly distinguished from the living systems by the values of R_T. Its values of R_L fall within the same range as those of the living systems. Of course many other features of the upper atmosphere of Titan differ from those of the living systems. We regard it as a strength of the present method of analysis that these dimensionless measures of disequilibrium permit a dimensionless quantitative comparison of such dissimilar systems with regard to this feature.

Note that the copolymer systems are far from equilibrium by design: Most combinatorial combinations are excluded by the preparation procedure which assures that the molecules will all consist of a string of one type of monomer attached by one bond to a string of the other type of monomer whereas the corresponding approximate

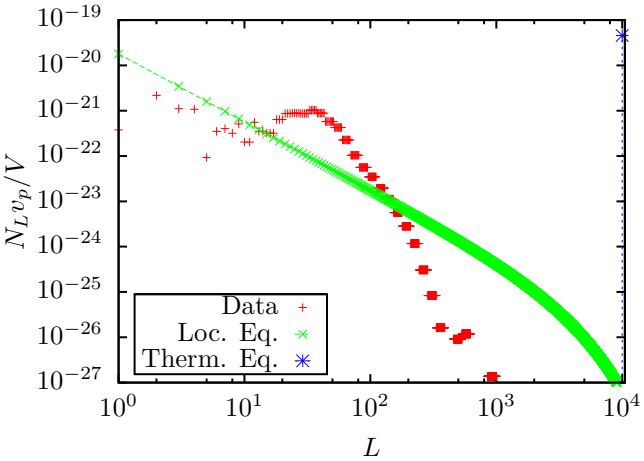

FIG. 13: $N_L v_p/V$ values for the Titan atmosphere at 1013km above the surface and corresponding equilibrium distributions.. $R_L = 0.29$ and $R_T = 0.71$.

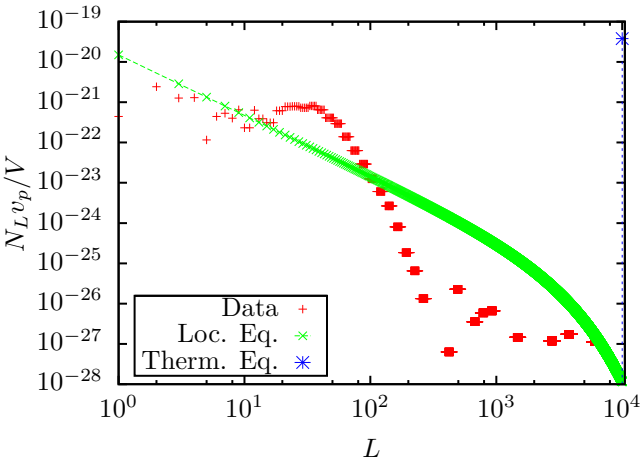

FIG. 14: $N_L v_p/V$ values for the Titan atmosphere at 1032km above the surface and corresponding equilibrium distributions. $R_L = 0.29$ and $R_T = 0.71$.

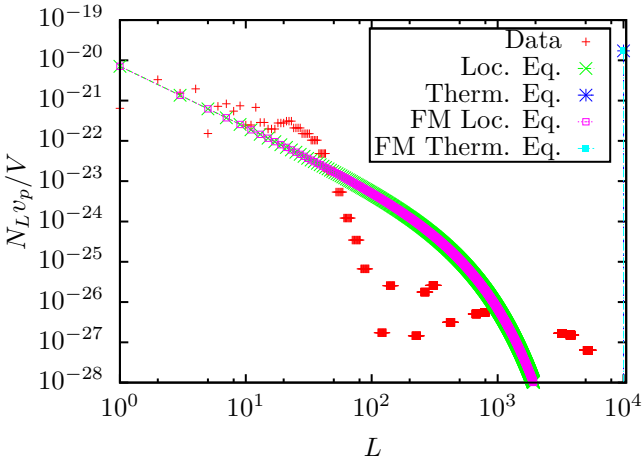

FIG. 15: $N_L v_p/V$ values for the Titan atmosphere at 1078km above the surface and corresponding equilibrium distributions. $R_L = 0.28$ and $R_T = 0.73$. The curves labeled 'FM Loc. Eq.' and 'FM Therm. Eq.' are the local and thermal equilibrium, respectively, calculated using the full model as described in [9].

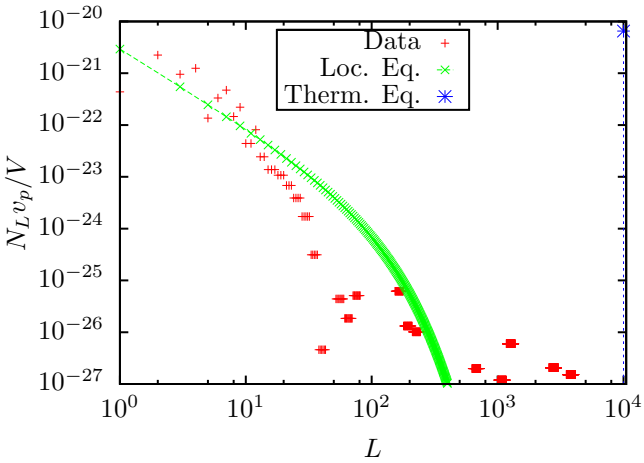

FIG. 16: $N_L v_p/V$ values for the Titan atmosphere at 1148km above the surface and corresponding equilibrium distributions. $R_L = 0.32$ and $R_T = 0.77$.

equilibria which we find take account of the possibility of mixing the two types of monomers in all possible ways along the chain.

All the molecules detected in the Titan atmosphere are almost certainly not linear as we have assumed in calculating the equilibrium distributions. Bonding rules assure that the energies will not be greatly affected by that error, but the degeneracies G_L which describe the numbers of molecules containing L monomers may be, particularly for small L. If as we anticipate, the error is larger for smaller L then this will shift the equilibrium distributions toward small L and it appears that this might increase the values of R_L. The analysis could be repeated with more detailed models of the small molecule chemistry of the Titan atmosphere, but the idea here was to use only the empirical data available, and that data does not currently provide the information on the distribution of molecular types which is required.

There are some issues with the R_T values for the Titan atmosphere: We used a reported value [12] for the temperature of the Titan atmosphere in the analysis. But that is not really consistent because we are finding that the Titan atmosphere is not in equilibrium and therefore cannot itself be characterized by a temperature. Secondly the temperature used for the Titan atmosphere in the determination of R_T is much lower ($T = 120K$) than the characteristic terrestrial temperature ($T = 293K$) used in the corresponding analysis of the living systems. However we have recalculated R_T for all the Titan data

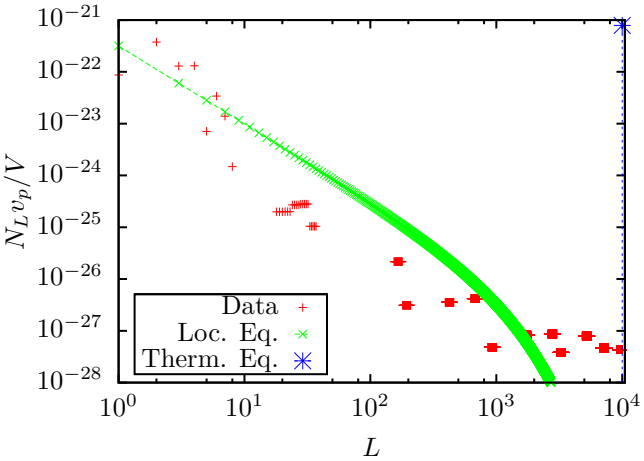

FIG. 17: $N_L v_p/V$ values for the Titan atmosphere at 1244km above the surface and corresponding equilibrium distributions. $R_L = 0.33$ and $R_T = 0.81$.

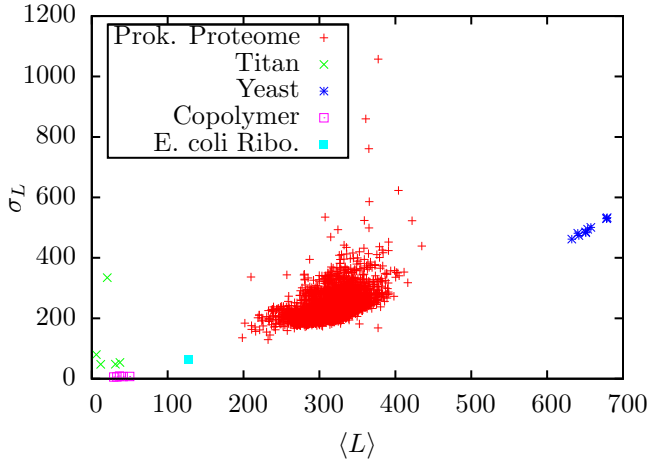

FIG. 18: Average length versus standard deviation of the length (σ_L).

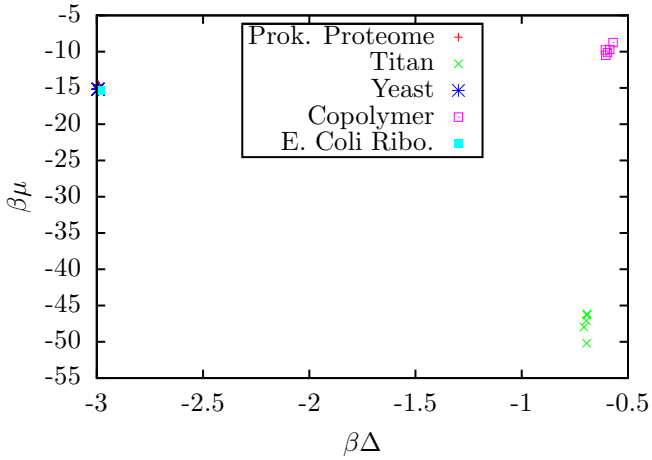

FIG. 19: Scatter plot of $\beta\mu$ and $\beta\Delta$ for the various systems.

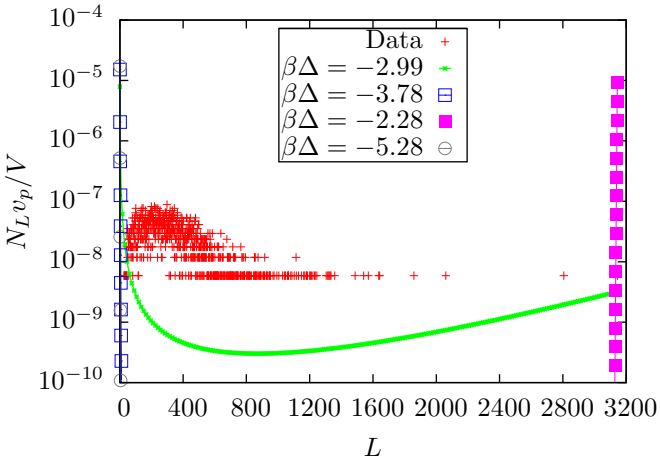

FIG. 20: The data of figure 6 compared to equilibria at different values of $\beta\Delta$. $\beta\Delta = -2.99$ corresponds to the local equilibrated distribution, $\beta\Delta = -3.78$ is the thermally equilibrated distribution using a bond energy of -2.2 kcal/mol and a temperature of 293K, $\beta\Delta = -2.28$ is the distribution using a bond energy of $(-2.2 + 0.875)$ kcal/mol and a temperature of 293K, and $\beta\Delta = -5.28$ is the distribution using a energy of $(-2.2 - 0.875)$ kcal/mol and a temperature of 293K. Here -2.2 kcal/mol and 0.875 kcal/mol are the average and standard deviation respectively of the protein bond energies found in [8].

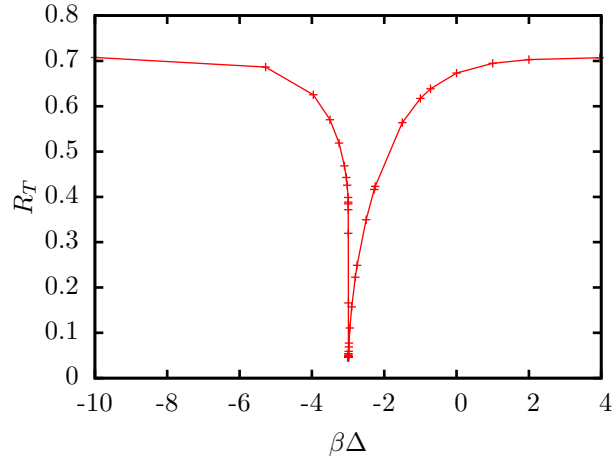

FIG. 21: R_T as a function of $\beta\Delta$ for a fixed data set.

assuming an ambient temperature of 293 K and find negligible changes.

We note that the molecular density of the organic molecules in the Titan upper atmosphere, both in absolute terms (particles per unit volume) and in dimensionless form as reported in figures 13-17, is much less than that of the living systems. One point which we are making in this work is that we can nevertheless compare its degree of equilibration with that for the living systems by a common measure. It has been a primary goal

of our program to find such measures, which can be used to analyse and compare systems found elsewhere to determine whether they have lifelike properties without imposing excessive terrocentric bias. We do not have data on the atmosphere of Titan near its surface, where the pressure exceeds that of the earth's surface atmosphere. However, as discussed briefly below, we do not anticipate the chemistry on the surface to be as far from equilibrium (and thus to have such large values of R_T and R_L) as that in the upper atmosphere, unless of course there is actually a biosphere on or below the solid surface of Titan.

The selection of the Titan data as a test case for applying these measures is based in part on our earlier work on Kauffman models, which suggest that very dilute chemical systems might have a better chance than denser ones of not falling into chemical equilibrium during their temporal evolution and this appears to be a minimal requirement for the spontaneous development of lifelike molecular dynamics. It has been known for a long time [22] that chemistry in the dilute environment of space, both in the upper atmosphere of earth and other planets and in the interplanetary medium, does in fact avoid falling into chemical equilibrium while sustaining substantial chemical activity. That condition is much closer to that of the small p Kauffman models that we previously simulated and which led to nonequilibrium, possibly lifelike, dynamics, than to the environments of ponds and oceanic trenches which are often speculated to be possible sites of prebiotic evolution. Astrophysical energy sources for sustaining prebiotic chemistry in such environments have been extensively studied (eg)[23]) and many of the chemical constituents believed to be required are present (eg [4, 24]).

Acknowledgments

This work was supported by the United States National Aeronautics and Space Administration (NASA) through grant NNX14AQ05G and used the used the computational resources of the Minnesota Supercomputing Institute, the Open Science Grid, the University of Minnesota School of Physics and Astronomy Condor cluster, and the NASA Advanced Supercomputing division Pleiades supercomputer. We thank Aaron Wynveen for helpful discussions and Ravindra Desai, Joao Paulo, Bastiaan Staal and Gabriel Vivo Truyols for answering questions about their work and supplying us with their data. The Titan data is available on NASA's Planetary Database System as well, in summary form in [2]

Appendix A: Conversion of Titan data

The reported data is of the form $(\log L_i, y_{L_i})$, where y_{L_i} is the number of counts in bin L_i centered at $\log L_i$ and L is the number of carbon or nitrogen atoms in a molecule. (The Cassini mass spectrometer did not have sufficient resolution to distinguish carbon from nitrogen masses.) To convert the reported distributions to a distribution as a function of L we suppose that the number y_{L_i}in the L_i^{th} bin represents a uniform distribution over the range $(1/2)(\log(L_{i-1}) + \log(L_i))$ and $(1/2)(\log(L_{i+1}) + \log(L_i))$. On a linear scale the corresponding range for L is from $\sqrt{L_i L_{i-1}}$ to $\sqrt{L_i L_{i+1}}$. We round the lower limit to the nearest larger integer and the upper limit to the nearest lower integer and attribute a count value of $y_{L_i}/(\sqrt{L_i L_{i+1}} - \sqrt{L_i L_{i-1}} + 1)$ to each integer within the interval.

Appendix B: Bond energies for Titan data

Because the N-N bond energy is much lower than the C-N and C-C bond energy as reviewed in Table II, the model for the energy expressed in the relation $E = -\sum_{L=1}^{l_{max}} (L-1) N_L \Delta$ cannot be used and the calculation of the equilibrium distributions involves, in principle, a more complicated statistical mechanical calculation which is described elsewhere[9]. Here we describe a kind of mean field approximation in which we use an average bond strength $\overline{\Delta}(p_c)$ in the expression for the energy: $E = -\sum_{L=1}^{l_{max}} (L-1) N_L \overline{\Delta}(p_c)$ where p_c is the fraction of the monomers in the system which are carbon. The average value which we use is

$$\overline{\Delta}(p_c) = \Delta_1 p_c^2 + 2\Delta_1 p_c(1-p_c) + \Delta_2(1-p_c)^2. \quad \text{(B1)}$$

where Δ_1 is the bond energy of the C-C and C-N bonds (assumed to be the same) and Δ_2 is the bond energy of the N-N bond. The expression implicitly assumed that the probability of finding a carbon at a site is p_c independent of its environment and that is not expected to be a very good assumption. To test the effects of the error, we consider the maximum and minimum energies per bond which a polymer of long length L could have given p_c and show that the results would not be much affected by using these extremal values.

To find the average bond strength which gives the minimum energy consider a polymer of length L and $L-1$ total bonds and that $L_c = p_c L$ of the monomers are carbon. The answer depends on whether $L_c \leq L/2$ or not. In the former case, the total energy is minimized by starting with a nitrogen atom and then alternating between carbon and nitrogen until all of the carbon atoms are used giving a sequence $NCNCNCNCNCN \cdots N$. (The NCNCNCNCNCN sequence can be placed anywhere in the chain, giving a degeneracy, but we do not consider that here.) For each carbon atom there are two carbon-nitrogen bonds, the rest are nitrogen-nitrogen bonds. This gives a total energy of $-2\Delta_1 L_c - \Delta_2(L-1-2L_c)$. For the other case $L_c > L/2$ start with all C atoms replace less than half of them with N atoms. This can be done without introducing any N-N bonds and, since we assume that the C-N and C-C bond energies are equal

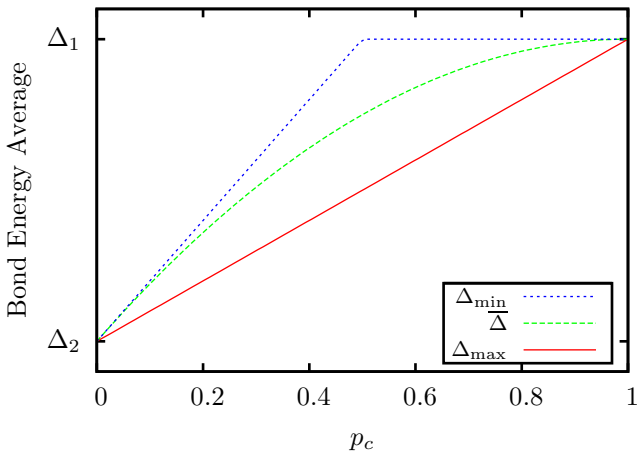

FIG. 22: The three bond energies defined in Appendix B (equations (B1), (B2),(B3)) as a function of the carbon proportion p_c.

we have a total minimum energy of $-(L-1)\Delta_1$. Dividing these energies by L and taking the large L limit with $p_c = L_c/L$ fixed we obtain an average bond energy giving a minimum polymer energy of

$$\Delta_{\min}(p_c) = \begin{cases} 2\Delta_1 p_c + \Delta_2(1-2p_c) & , \ p_c \leq 0.5 \\ \Delta_1 & , \ p_c > 0. \end{cases} . \quad \text{(B2)}$$

The calculation of the average bond energy which maximizes the total energy ($\Delta_{\max}$) given L atoms is simpler: One must maximize the number of N-N bonds and that is done by connecting all the carbon atoms together, connecting all the nitrogen together, and then connecting the two with a single carbon-nitrogen bond ($C\cdots CCNN\cdots N$). The total energy is given by $-\Delta_1 L_c - \Delta_2(L-1-L_c)$. Dividing this total energy by L and taking the limit $L \to \infty$ yields the average bond energy which maximizes the total energy:

$$\Delta_{\max}(p_c) = \Delta_1 p_c + \Delta_2(1-p_c). \quad \text{(B3)}$$

The three bond energies, $\overline{\Delta}$, $\Delta_{\min}$, and $\Delta_{\max}$ are plotted in figure 22. Using each of these three bond energies for all the bonds we computed the values of $\beta\mu$,$\beta\Delta$ for the Titan data for some test cases and found that the results were not sensitive to the value used. Using the three bond energies $\overline{\Delta}$, $\Delta_{\min}$, and $\Delta_{\max}$ for calculating the thermal equilibrium length distribution for the Titan 1078km data it was found that the value of $\beta\mu$ changed by only 1% and there was no discernible change in the value of R_T. We also have preliminary, results plotted in 15, for the full model as described in [9] where one sees that the full model gives results quite close to those found by the approximate approach described above and used to obtain the results in the rest of the Titan data considered here.

[1] B. F. Intoy and J. W. Halley, , Phys. Rev. E 96, 062402 (2017)

[2] R. Desai, A. Coates, A. Wellbrock, V. Vuitton, F. Crary, D. González-Caniulef, O. Shebanits, G. Jones, G. Lewis, J. Waite, et al. Carbon Chain Anions and the Growth of Complex Organic Molecules in Titans Ionosphere. *The Astrophysical Journal Letters*, 844(2):L18, 2017.

[3] J. H. Waite Jr., D. T. Young, T. E. Cravens,A. J. Coates, F. J. Crary B. Magee, J. Westlake, Science 316, 870 (2007)

[4] Martin Rahm, Jonathan I. Lunine, David A. Usher, and David Shalloway,Proceedings of the National Academy of Sciences of the United States of America 113(29), pp. 8121-8126 (2016)

[5] A. Wynveen, I. Fedorov, and J. W. Halley, Phys. Rev. E 89, 022725 (2014)

[6] B. F. Intoy, A. Wynveen and J. W. Halley, Phys. Rev. E 94, 042424 (2016)

[7] Jonathan E. Kohn, Ian S. Millett, Jaby Jacob, Bojan Zagrovic, Thomas M. Dillon, Nikolina Cingel, Robin S. Dothager, Soenke Seifert, P. Thiyagarajan, Tobin R. Sosnick, M. Zahid Hasan, Vijay S. Pande, Ingo Ruczinski, Sebastian Doniach, and Kevin W. Plaxco Random-coil Behavior and the Dimensions of Chemically Unfolded Proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34):12491–12496, 2004. (Erratum: PNAS 102 (40) 14475 (2005)).

[8] R. B. Martin. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers*, 45(5):351–353, 1998.

[9] in preparation.

[10] P Valerino and B Buffington, Titan Event Summary Table. Reference Trajectory 110818v4, NASA Planetary Data System, 2013.

[11] Uniprot: The Universal Protein Knowledgebase. *Nucleic acids research*, 45(D1):D158–D169, 2017.

[12] F. Crary, B. Magee, K. Mandt, J. Waite Jr, J. Westlake, and D. Young. *Planetary and Space Science*, 57(14-15):1847–1856, 2009.

[13] T. Cravens, R. Yelle, J.-E. Wahlund, D. Shemansky, and A. Nagy. Composition and Structure of the Ionosphere and Thermosphere. In *Titan From Cassini-Huygens*, pages 259–295. Springer, 2009.

[14] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[15] R. Milo. What is the Total Number of Protein Molecules per Cell Volume? a Call to Rethink Some Published Values. *Bioessays*, 35(12):1050–1055, 2013.

[16] J. A. Paulo, J. D. O'Connell, R. A. Everley, J. O'Brien, M. A. Gygi, and S. P. Gygi. Quantitative Mass Spectrometry-based Multiplexing Compares the Abundance of 5000 *S. cerevisiae* Proteins Across 10 Carbon Sources. *Journal of proteomics*, 148:85–93, 2016.

[17] G. Vivó-Truyols, B. Staal, and P. J. Schoenmakers. Strip-based Regression: A Method to Obtain Comprehensive Co-polymer Architectures from Matrix-assisted Laser Desorption Ionisation-mass Spectrometry Data.

Journal of Chromatography A, 1217(25):4150–4159, 2010.

[18] B. B. P. Staal. *Characterization of (co) polymers by MALDI-TOF-MS*. Technische Universiteit Eindhoven Eindhoven, The Netherlands, 2005.

[19] S. S. Zumdahl and S. A. Zumdahl. *Chemistry*, chapter 8, page 351. Houghton Mifflin Company, Boston, MA, 7th edition, 2007. Table 8.4.

[20] N. Fischer,P. Neumann,A. L. Kovevega,L. V. Bock,R. Ficner, M. V. Rodnina,H. Stark, H. , Nature 520: 567 (2015)

[21] R. Milo and R. Phillips. *Cell biology by the numbers*, chapter 2. Garland Science, 2015.

[22] J.P Ferris, L. Becker, K. Boering, G.D. Cody, G.B. Ellison, J. M Hayes, R. E. Johnson, W. Klemperer, K. J. Meech, K. S. Noll and M. Saunders, Exploring Organic Environments in the Solar System, The National Academies Press, Washington D. C. (2007)

[23] Manasvi Lingam, Chuanfei Dong, Xiaohua Fang, Bruce M. Jakosky and Abraham Loeb The Astrophysical Journal, 853,10a (2018)

[24] L. Majumdar et al, Monthly Notices of the Royal Astronomical Society: Letters 473, L59-L63 (2018)

[25] P. C. Hiemenz and T. P. Lodge. *Polymer chemistry*. CRC press, 2007. Table 6.1.